

Generalized Probabilistic Theories Without the No-Restriction Hypothesis

Peter Janotta¹ and Raymond Lal²

¹Universität Würzburg, Am Hubland, Fakultät für Physik und Astronomie, 97074 Würzburg, Germany

²University of Oxford, Department of Computer Science,
Quantum Group, Wolfson Building, Parks Road, Oxford OX1 3QD, UK.

The framework of generalized probabilistic theories (GPTs) is a popular approach for studying the physical foundations of quantum theory. The standard framework assumes the no-restriction hypothesis, in which the state space of a physical theory determines the set of measurements. However, this assumption is not physically motivated. We generalize the framework to account for systems that do not obey the no-restriction hypothesis. We then show how our framework can be used to describe new classes of probabilistic theories, for example those which include intrinsic noise. Relaxing the restriction hypothesis also allows us to introduce a ‘self-dualization’ procedure, which yields a new class of theories that share many features of quantum theory, such as obeying Tsirelson’s bound for the maximally entangled state. We then characterize joint states, generalizing the maximal tensor product. We show how this new tensor product can be used to describe the convex closure of the Spekkens toy theory, and in doing so we obtain an analysis of why it is local in terms of the geometry of its state space. We show that the unrestricted version of the Spekkens toy theory is the theory known as ‘boxworld’ that allows maximal nonlocal correlations.

I. INTRODUCTION

The framework of generalized probabilistic theories (GPTs) is a modern operational approach for studying the physical foundations of quantum theory [1]. The framework is *operational* because a theory is defined according to the observable measurement statistics that it predicts. In contrast, quantum theory is usually defined using an abstract mathematical formalism without physical motivation (e.g. the density matrix formalism). Assuming only basic principles, the framework encompasses a large variety of theories. For example, quantum theory and classical probability theory are special cases of GPTs. The focus of work on GPTs is to identify the unique physical properties that distinguish quantum theory from other theories. More generally, one can examine the relationship between different physical properties, such as no-cloning and nonlocality, without restricting to a particular physical theory.

Using this framework, it has been shown that many properties that were thought to be particular to quantum theory are in fact very general. As a sample of such results, it was shown that any non-classical probability theory (in the sense to be described in section II) has the following properties: the existence of entanglement [1]; for mixed states, the lack a unique decomposition into a unique ensemble of pure states; generalizations of the no-cloning or no-broadcasting theorem [2]; and, an information-disturbance trade-off [3]. Notably, recent attempts to reconstruct quantum theory from physical axioms include the assumptions made in GPTs [4] or very similar assumptions [5, 6].

A GPT is defined by a set of preparations, a set of measurements, and composition rules for multipartite systems called the tensor product of the theory. In general there is a trade-off between possible preparations and possible measurement outcomes: the larger the set of preparations, the smaller the upper bound on the set of allowed measurements [7]. In the existing GPT framework, it is usually assumed that this upper bound is *saturated*. This means that, for a chosen set of states, *all* potential measurement outcomes that yield probability-valued

results are assumed to be physically realizable. This is called the *no-restriction hypothesis* [6]. This assumption is not based on any physical motivation, and it is usually assumed for the sake of mathematical convenience. In this work we take on the task of extending the framework of GPTs when the no-restriction hypothesis is abandoned. This extension of GPTs therefore brings the framework closer to the operational motivation for which it was originally initiated.

Our contribution. The idea of removing the no-restriction hypothesis (or replacing it with other assumptions) has appeared sporadically in other works [6, 8]. However, until now a systematic analysis of the consequences of doing so has been lacking. In this paper we provide a well-defined framework with the no-restriction hypothesis omitted, whilst keeping the other assumptions of the GPT framework. Our work then proceeds in two parts.

In the first part we show that this new framework encompasses more theories than before. For example, we show that theories with intrinsic noise can be described in our framework, but not in the existing GPT framework. We also provide a procedure for constructing a *self-dual* theory from a theory which is not self-dual. The importance of this is that self-duality has been shown to imply ‘quantum-like’ (for example, limiting bipartite nonlocality to Tsirelson’s bound for the maximally entangled state [9]). Hence this allows us to introduce a new class of probabilistic theories with ‘quantum-like’ behaviour, and crucially, this is a class of theories which does not satisfy the no-restriction hypothesis.

In the second part, we develop the treatment of composite systems. In particular, we show that our extension requires a new (and more general) definition of the tensor product for describing composite systems. This significantly extends the GPT framework, since it allows us to analyse the relationship between nonlocality and the geometry of the state space of a theory, building on previous work in this direction. For example, we show how the Spekkens toy theory (for which the connection to GPTs had not been previously established) can be viewed as a GPT, but only in our more general framework. Moreover, this allows us to give an analysis of why the

Spekkens theory is local, using the geometry of its state space.

Structure of the paper. In section II we give a brief overview of the framework of GPTs. We then begin the first part of our analysis, concentrating on single systems. In section III we describe in detail the no-restriction hypothesis, and some consequences of relaxing this assumption. In section IV we develop the important example of theories with noise. In section V we introduce the self-dualization procedure, and discuss the class of theories that this introduces. We then enter the second part of our analysis, which concerns composite systems. In section VI we explain how joint states of composite systems are usually described. In section VII we show why a new definition of composite systems is needed, and we introduce this definition. We then study examples of theories such as the Spekkens model.

II. GENERALIZED PROBABILISTIC THEORIES: A BRIEF SUMMARY

A physical experiment consists of the following steps: the preparation of a system, transformations of that system (e.g. by inherent dynamics), and a measurement. In general, the measurement will different outcomes, each occurring with some probability. Defining a generalized probabilistic theory amounts to specifying these probabilities for any such combination of preparation, transformation and measurement. Note that transformations can be absorbed into either the preparation or the measurement. Hence to define the allowed probability distributions of a GPT, it suffices to define the set of preparation procedures and the set of measurements.

1. States and effects

Consider a class of preparation procedures which all yield exactly the same measurement statistics. The members of this class are experimentally indistinguishable. Since a GPT concerns only experimental statistics, we can define a *state* of a system as such an equivalence class. Analogously we also define an *effect* as an equivalence class of measurement outcomes. We will refer to this identification of states and effects with their respective measurement statistics as the *equivalence principle*. Mathematically, states are represented by elements of a vector space V . Effects are linear functionals on states, i.e. elements of the dual space V^* . Applying an effect e to a state ω yields the probability $p(e|\omega) = e(\omega)$ for the corresponding measurement outcome to occur when measuring the system in the state. Without loss of generality we will choose a specific representation of states and effects in this paper to demonstrate the abstract concepts. Both states and effects will be represented by vectors embedded in \mathbb{R}^n . The application of effects on states is given by the Euclidean inner product of the respective vectors:

$$e = (\varepsilon_1, \dots, \varepsilon_n)^T \quad \omega = (w_1, \dots, w_n)^T \quad (1)$$

$$p(e|\omega) = e^T \cdot \omega = \sum_i \varepsilon_i w_i \quad (2)$$

The GPT framework also accounts for ensembles of preparations or measurements, in which there is uncertainty about which measurement is implemented, or which state has been prepared. This could occur if there is a probabilistic selection of the preparation procedure, for example. This probability distribution is represented by using *mixed states* and *mixed effects*, given by convex combinations:

$$e = \sum_i \lambda_i e_i \quad \lambda_i \geq 0, \sum_i \lambda_i = 1 \quad (3)$$

$$\omega = \sum_i \mu_i \omega_i \quad \mu_i \geq 0, \sum_i \mu_i = 1 \quad (4)$$

corresponding to ensembles $\{\lambda_i, e_i\}$ and $\{\mu_i, \omega_i\}$. Consequently, states and effects form convex sets. If the only convex decomposition of a state ω is such that $\omega \propto \omega_i$ for all i , then the state is a *pure state*. Similarly, if the only convex decomposition of an effect e using Eq. 1 is such that $e \propto e_i$ for all i , then the effect is a *pure effect*.

Since effects and states act linearly on each other, the probability distribution for the ensembles is the weighted sum of the probabilities $p_{ij} = e_i(\omega_j)$ of individual ensemble elements:

$$e(\omega) = \sum_{i,j} \lambda_i \mu_j e_i(\omega_j). \quad (5)$$

More generally, consider the result of applying different measurements to systems prepared by the same method. In general, there will be measurement outcomes with probabilities that are linearly dependent for a fixed state. Analogously, one might find linear dependencies between the probabilities for a fixed measurement outcome under variations of the state that is prepared. This implies a linear dependence between the vector space elements $\omega \in V$ representing the states; there is a corresponding linear dependence for the effects $e \in V^*$. This determines the *dimension* of V as the minimal number of different measurement outcomes needed to identify a state uniquely (this is called the ‘fiducial set’ of measurement outcomes by Hardy [10]). In this paper we restrict ourselves to systems for which the vector space V has finite dimension. Hence the dimension of V is equal to the dimension of the dual space V^* , which is the minimal number of preparations required to identify an effect.

2. Normalization and measurements

A central concept in the GPT framework is the description of *perfect* preparations and measurements. A perfect preparation is one that is guaranteed to succeed. It is represented by a normalized state, where normalization defined with respect to a special effect, called the *unit measure* u . The set of all normalized states is called the *state space* Ω . The unit measure u represents an unbiased measurement with only one outcome: this outcome occurs if a preparation has succeeded, i.e. it is determined by

$$u(\omega) = 1 \quad \forall \omega \in \Omega. \quad (6)$$

In the specific representation used in this paper we choose

$$u := (0, \dots, 0, 1)^T.$$

Consequently, for a state ω embedded in an n -dimensional vector space V , the normalization of ω is directly apparent from the last component ω_n , i.e. normalized states have $\omega_n = 1$.

An effect is a map $e : \Omega \rightarrow [0, 1]$ that gives a probability when applied to a normalized state ω . A perfect measurement consists of a set of effects $\{e_i\}$ which sum up to the unit measure, i.e.:

$$\sum_i e_i = u.$$

Thus, measurement probabilities sum up to one for any perfectly-prepared system.

Beyond the description of perfect preparations and measurements, the GPT framework also accounts for the opposite extreme, namely preparations that always fail or measurement outcomes that never occur no matter which state they are applied to. The corresponding states and effects are given by the zero elements \emptyset of V and V^* with

$$\emptyset(\omega) = 0 \quad \forall \omega \in V \quad (7)$$

$$e(\emptyset) = 0 \quad \forall e \in V^*. \quad (8)$$

Imperfections in preparations yield unnormalized states resulting from the mixture of a normalized state ω and \emptyset . Detector deficiencies and bias can be addressed by mixing every effect of a perfect measurement with \emptyset or another common effect. However, we will show in section VI 3 that consistency conditions on joint states forbid imperfect measurements. Consequently, the measurement has to be completed by an additional effect, such that the effects sum up to the unit measure, even though the occurrence of this additional measurement outcome cannot be registered by an experimenter due to detector deficiencies.

3. Equivalent Representations

Consider applying arbitrary bijective linear maps L^T on all effects and the corresponding inverse map L^{-1} on all states. This leaves the results from any combination of effects and states invariant, since:

$$(L^T \cdot e)[L^{-1} \cdot \omega] = (L^T \cdot e)^T \cdot L^{-1} \cdot \omega = e^T \cdot L \cdot L^{-1} \cdot \omega = e^T \cdot \omega. \quad (9)$$

Now, a particular probabilistic theory is associated with a particular state space Ω and set of effects E . But theories are distinguished only by the different measurement statistics that are possible (as is guaranteed by using the equivalence principle). Hence if Ω and E are transformed according to (9), then the resulting Ω' and E' define the same theory, since this transformed state space and effect set yield the same measurement statistics.

4. Examples

Quantum theory. Consider the usual quantum formalism, for which a state is given by a density matrix ρ on a Hilbert space \mathcal{H} . By decomposing density matrices in an operator basis, we obtain the real vector space V defined above for quantum theory. For example, there is a well-known representation of the normalized states of a qubit as a linear combination of the Pauli-operators σ_i :

$$\rho = \frac{1}{2}(\mathbb{1} + a\sigma_x + b\sigma_y + c\sigma_z) \quad a^2 + b^2 + c^2 \leq 1 \quad (10)$$

Forming a real vector from the coefficient a, b, c gives the representation of the qubit state space in $V = \mathbb{R}^3$: this is the Bloch ball.

Adding a fourth component that indicates normalization gives a representation similar to (1). However, for quantum systems of higher dimension the characterization of the geometrical shape of the state spaces in this representation is still an open problem [11].

In the usual density matrix representation an effect is a POVM element E , which is applied via the trace rule, so that the probability of an effect E given the state ρ is given by $\text{Tr}[E \circ \rho]$. The unit measure u is given by the identity operator $\mathbb{1}$ on \mathcal{H} , so that a density matrix ρ is normalized when:

$$\text{Tr}[\mathbb{1} \circ \rho] = 1.$$

Note that for quantum systems the set of states and the set of effects can be *identified*: this is the set of positive operators on \mathcal{H} . For example, for a qubit the Bloch ball represents both (normalised) states and effects. This is an example of ‘self-duality’ in a theory; we shall discuss this further in section V.

Classical probability theory. The state space of a classical system in \mathbb{R}^d is a simplex. This is the convex hull of $d + 1$ pure states (which can be characterized via a condition on linear independence). For example, for $d = 1$, the classical state space is a geometrically line, which represents a bit. The extreme points of the line ω_0 and ω_1 are the pure states: these represent the values 0 or 1 of the bit respectively. The convex mixtures $p\omega_0 + (1 - p)\omega_1$ represent states of classical uncertainty about the value of the bit. Only one measurement outcome is needed to identify the state, e.g. the probability of obtaining the 0 value for the bit. For $d = 2$, the simplex is a triangle in \mathbb{R}^2 , which represents a trit; and so on. As for a bit, for any d the pure states ω_i represent mutually exclusive properties of the system. For example, if one knows with certainty which number is on top of a die, then one automatically knows that none of the other numbers is on top. This means that the pure effects then correspond to measurement outcomes that perfectly distinguish ω_i i.e. $e_i(\omega_j) = \delta_{ij}$.

Boxworld. This is a popular toy theory in the GPT framework that is neither quantum nor classical, which was first introduced systematically in [1]. Boxworld consists of a class of single systems characterized by the dimension $d \geq 2$ of the state space. For $d = 2$ the normalized state space Ω is the

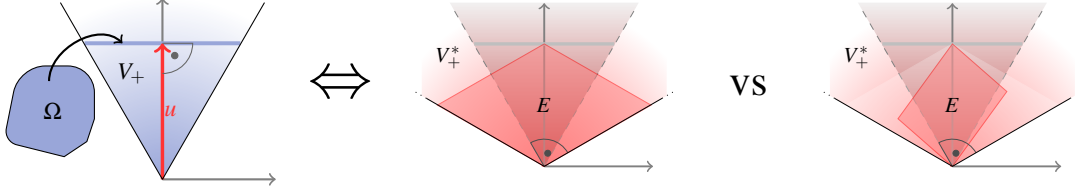


FIG. 1. The construction of the effect set E in the traditional GPT framework with no-restriction hypothesis is shown in the middle. Without the no-restriction hypothesis the definition of the effect set gets an independent part of the theory specification (right picture).

convex hull of the following pure states:

$$\omega_1 = (1, 0, 1)^T \quad \omega_2 = (0, 1, 1)^T \quad (11)$$

$$\omega_3 = (-1, 0, 1)^T \quad \omega_4 = (0, -1, 1)^T, \quad (12)$$

and so geometrically Ω is a square. The set of effects is given by the convex hull of $\emptyset = (0, 0, 0)^T$, $u = (0, 0, 1)^T$ and the following extremal effects:

$$e_1 = \frac{1}{2}(1, 1, 1)^T \quad e_2 = \frac{1}{2}(-1, 1, 1)^T \quad (13)$$

$$e_3 = \frac{1}{2}(-1, -1, 1)^T \quad e_4 = \frac{1}{2}(1, -1, 1)^T \quad (14)$$

It is straightforward to show that the measurement statistics of the two orthogonal binary measurements $M_1 = \{e_1, e_3\}$ and $M_2 = \{e_2, e_4\}$ give enough information to identify any state. Indeed, due to the normalization constraint the measurement statistics of the binary measurements on normalized states is determined by the probabilities p_1, p_2 for the first outcomes e_1, e_2 . The different states give rise to the full range $(p_1, p_2) \in [0, 1]^2$ of possible probability distributions, with the probabilities p_1 and p_2 being independent. Hence the measurement outcomes e_1 and e_2 are enough to identify the state of the system, which verifies that the dimension is $d = 2$. Note that unlike orthogonal measurements in quantum theory (such as σ_x and σ_y), there is no uncertainty principle for M_1 and M_2 for this system [12]. For example, although e_1 and e_4 belong to orthogonal measurements, we have $e_1(\omega_1) = e_4(\omega_1) = 1$.

Higher dimensional single systems with $d > 2$ in boxworld have d different binary orthogonal measurements and state spaces given by hypercubes. For the joint states that we shall discuss in section VI, boxworld allows maximal nonlocal correlations (using the CHSH inequality introduced below). These correlations define the *Popescu-Rohrlich box* [13], and they are not realizable by quantum theory.

III. THE NO-RESTRICTION HYPOTHESIS

We now consider in detail the no-restriction hypothesis, and the consequences of relaxing it.

A. Defining the set of effects

Effects are restricted to give values in the range of $[0, 1]$ when applied to normalized states. But in the traditional

framework of GPTs, the set of effects E is not restricted any further. That is, the set of effects is exactly the set of all probability-valued linear functionals on the given states. We will call this relationship between states and effects the *no-restriction hypothesis*, in accordance with [6]. It is satisfied for classical probability theory and quantum theory.

Theorem 1. *The set of effects under the no-restriction hypothesis is given by*

$$E := V_+^* \cap (u - V_+^*) \quad (15)$$

with the so-called dual cone

$$V_+^* := \{e \in V^* \mid e(\omega) \geq 0 \quad \forall \omega \in \Omega\}. \quad (16)$$

Proof. The definition of effects as probability-valued linear functionals can be decomposed into two conditions.

The first condition is that effects have to give non-negative results on every element of the state space. For arbitrary elements $e \in V^*$ satisfying this condition, the condition is also satisfied by the positive ray $\{\lambda e \mid \lambda \geq 0\}$. Hence, the set obeying the non-negativity condition is a cone, namely the *dual cone* V_+^* , defined by (16).

The second condition on effects requires them to give results not larger than one, when applied to arbitrary normalized states. In other words the results have to be one or one minus a positive value, i.e. $e \in u - V_+^*$.

In the standard framework both boundary conditions are saturated. That is, for a given state space, *any* linear functional that gives probability-valued results for all normalized states is included in the theory. Thus, the set of effects E is $V_+^* \cap (u - V_+^*)$. \square

The dual of the dual cone is the *primal cone* V_+ , which is generated by unnormalized states, i.e.

$$V_+ := \{\lambda \omega \mid \omega \in \Omega, \lambda \geq 0\} = (V_+^*)^*. \quad (17)$$

Consequently, if the no-restriction hypothesis holds, then a theory is completely determined by the state space, since the effect set can be derived from the state space.

The purpose of this paper is to develop the framework of GPTs without the no-restriction hypothesis. There are two main reasons for doing so. Firstly, the necessity of the no-restriction hypothesis is questionable from an operational perspective. Indeed, considering the physical meaning of states and effects there is no reason to believe that the possible preparation procedures determine possible measurements.

Secondly, this will generalize the GPT framework to cover new scenarios that have not been accessible within the old framework.

B. Relaxing the no-restriction hypothesis

Let us note the constraints that still apply when the no-restriction hypothesis is removed. Clearly, effects still need to give probabilities when applied to any state. That is, when allowing violations of the no-restriction hypothesis, the set of probability-valued linear functionals on states in (15) remains an upper bound for possible effects. However, in general not all elements in this set need to represent a valid measurement outcome. Consequently, the set of effects E may actually be given by a *subset* of (15). This is the crucial new ingredient in the GPT framework that we shall use in subsequent sections.

Furthermore, we have identified the following four consistency conditions that also have to be met:

- i) The unit measure u needs to be included in the restricted set as it is crucial for the definition of measurements.
- ii) For every effect e included in E , the complement effect $\bar{e} = u - e$ needs to be included as well. We will show in section VI 3 that including an effect, but not the complement can yield inconsistencies for joint states.
- iii) Coarse graining also provides effects that can be derived from existing ones. If one does not distinguish between some measurement outcomes that are part of the same measurement, the common probabilities are given by the sum of the individual probabilities. Due to linearity the corresponding effect describing the coarse graining is given by the sum of the individual effects.
- iv) Transformations map valid states to valid states. However, for any transformation T on states, there is also an adjoint transformation T^\dagger on effects defined by $e[T(\omega)] = [T^\dagger(e)](\omega)$ for all states and effects. Thus, the effect set has to respect given transformations.

Apart from these consistency restrictions, the definition of the effect set E is now an independent part of the specification of the theory. In other words, the effect set E does not depend on the state space now, and the dual cone V_+^* is irrelevant for single systems. However, we will see in section VII B that we still need it to classify consistent joint states.

Let us now consider how removing the no-restriction hypothesis will be useful. As shown above, the no-restriction hypothesis connects a set of states and effects via the respective dual-cone. Taking a closer look at the dual cone construction in (16), it can easily be seen that each extremal point of the primal cone describes a facet of the dual cone and the other way round. Therefore, arbitrary small changes in the primal cone, can have an enormous impact on the form of the dual cone. Consequently, the no-restriction hypothesis makes it extremely difficult to alter a theory in a controlled way. However, it has always been a central motivation for the framework of generalized probabilistic theories to find alternatives to quantum theory.

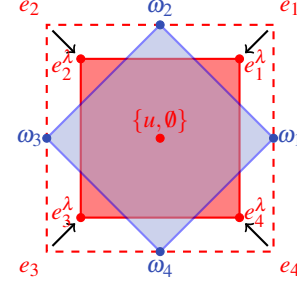


FIG. 2. Inclusion of noise into boxworld: State space and effects are both embedded into \mathbb{R}^3 and shown from above for illustration. The state space (blue) is given by a square. The effect set is the octahedron spanned by the extremal effects e_i , u and \emptyset . The noisy theory has a restricted effect set with extremal effects e_i^λ .

We shall now show in sections IV and V that new models with interesting features can indeed be constructed when accepting violations of the no-restriction hypothesis. Furthermore, for joint systems, we will see in sections VI and VII how consistency conditions are affected.

IV. THEORIES WITH INTRINSIC NOISE

The no-restriction hypothesis guarantees that for any pure state ω , there is an effect e , with $e \neq u$, such that $e(\omega) = 1$. In contrast, removing the no-restriction hypothesis allows for the modeling of systems with intrinsic noise, i.e. systems for which the unit measure is the *only* certain outcome for any state. For example, an isotropic unbiased implementation of noise can be achieved by restricting the effects to a set where the original extremal effects are replaced by mixtures with $u/2$ (except for \emptyset and u itself). In order to combine noise and bias one can mix the extremal with another effect instead of $u/2$.

The inclusion of intrinsic noise by a modification of boxworld is illustrated in Fig. 2. The state space of a single system is given by a square. In the traditional model the effect set is determined by the no-restriction hypothesis. A noisy version of boxworld is given by mixing the extremal effects e_i with $u/2$:

$$e_i \mapsto e_i^\lambda = \lambda e_i + (1 - \lambda) \frac{u}{2} \quad (18)$$

The strength of noise is given by $(1 - \lambda)$, i.e. the maximal probability from extremal effects is λ .

This model is particularly interesting with respect to its potential non-local correlations in joint systems. This will be examined in more detail after introducing joint states in section VI.

V. SELF-DUALIZATION PROCEDURE

A particular class of systems that has gained a lot of interest recently are so-called (*strongly*) *self-dual* systems [2, 9, 14]. These are systems with a particular geometrical structure,

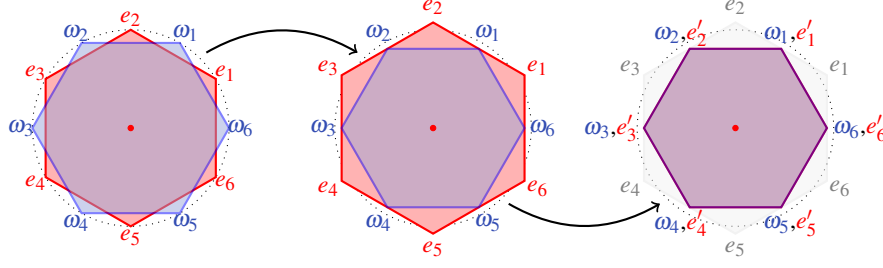


FIG. 3. Self-dualization of a hexagon system: The pictures show the statespace (blue) and the intersection of the effect cone (red) that lies in the same plane. In the first step the state cone will be embedded into the effect cone by an equivalence transformation (9). In the second step the effects not included in the state cone are abandoned.

shared by both classical probability theory and quantum theory. For strongly self-dual systems states and effects can be identified with each other and thus be represented by the same mathematical objects. E.g. in quantum theory both states and effects are represented by positive hermitian operators.

Formally, strong self-duality is given by the following definition.

Definition 2. A system is *strongly self-dual* iff there exists an isomorphism $\Phi : V_+^* \mapsto V_+$ giving rise to a corresponding symmetric bilinear form T with $T(e, f) = e[\Phi(f)] = T(f, e)$ and $T(e, e) \geq 0$ for all $e, f \in V^*$.

That is, T provides a semi inner product on effects. In a similar way for strongly self-dual systems the inverse map Φ^{-1} leads to a semi inner product on states.

Strong self-duality greatly restricts the class of possible systems. As we describe below, the property of ‘bit-symmetry’ implies that a system is strongly self-dual [14], and there is evidence that non-local correlations of self-dual systems are limited [9]. In this section we provide a general construction rule to modify any system, such that it resembles the behaviour of strongly self-dual systems.

Theorem 3. Any theory in the GPT framework can be modified to resemble strongly self-dual systems respecting Definition 2 with the dual cone V_+^* replaced by a truncated cone \mathcal{V}_+^* .

Proof. Using our representation, we assume an embedding of effects and states in a common vector space with a scalar product mediating the application of effects on states, as in Eq. 1. We start from an arbitrary theory for which the no-restriction hypothesis holds. The freedom of linearly transformations L^T from (9) allows us to strictly enlarge the effect cone V_+^* , while the corresponding inverse L^{-1} constricts the cone of unnormalized states V_+ to be strictly smaller. Hence, one can always represent the same physical theory, with V_+ embedded in V_+^* . We can then define a truncated the effect cone from $\mathcal{V}_+^* \subseteq V_+^*$, such that \mathcal{V}_+^* coincides with the state cone V_+ . Hence we can describe unnormalized effects and states with the same set of vectors. Consequently, the restriction of effects yields the vector space’s scalar product to act as an inner product between states. This satisfies the definition of strong self-duality with the dual cone V_+^* exchanged for the truncated

effect cone \mathcal{V}_+^* . The set of effects is then constructed from \mathcal{V}_+^* by $E = \mathcal{V}_+^* \cap u - \mathcal{V}_+^*$. \square

The connection between self-dualized systems and actual strongly self-dual systems is not only limited to a mere formal resemblance. In fact, the following example shows that self-dualized systems have features that strongly self-dual systems have when the no-restriction hypothesis is assumed.

A. Example: self-dualized polygons

Let us illustrate the self-dualization procedure on a set of systems introduced in a previous paper [9]. It is defined by two-dimensional state spaces with the shape of regular polygons. While the cases with an odd number of vertices n are strongly self-dual, the even cases are not.

For fixed n , let Ω be the convex hull of n pure states $\{\omega_i\}$, $i = 1, \dots, n$, with

$$\omega_i = \begin{pmatrix} r_n \cos(\frac{2\pi i}{n}) \\ r_n \sin(\frac{2\pi i}{n}) \\ 1 \end{pmatrix} \in \mathbb{R}^3, \quad (19)$$

where $r_n = \sqrt{\sec(\pi/n)}$.

The unit effect is

$$u = (0, 0, 1)^T. \quad (20)$$

The set $E(\Omega)$ of all possible measurement outcomes will be determined by the no-restriction hypothesis. In the case of even n , $E(\Omega)$ is the convex hull of the zero effect, the unit effect, and e_1, \dots, e_n , with

$$e_i = \frac{1}{2} \begin{pmatrix} r_n \cos(\frac{(2i-1)\pi}{n}) \\ r_n \sin(\frac{(2i-1)\pi}{n}) \\ 1 \end{pmatrix}. \quad (21)$$

The odd case yields a different expression for the ray-extremal effects

$$e_i = \frac{1}{1 + r_n^2} \begin{pmatrix} r_n \cos(\frac{2\pi i}{n}) \\ r_n \sin(\frac{2\pi i}{n}) \\ 1 \end{pmatrix}. \quad (22)$$

As shown in III the complement effects $\bar{e}_i = u - e_i$ of ray extremal effects e_i are also extremal in the effect set $E(\Omega)$. Whereas for even n these happen to coincide with $\bar{e}_i = e_{(i+n/2) \bmod n}$, for odd n the complement effects form additional extremal points of $E(\Omega)$. In summary, $E(\Omega)$ is the convex hull of the zero effect, the unit effect u , the ray-extremal effects e_1, \dots, e_n , and for odd n additionally $\bar{e}_1, \dots, \bar{e}_n$.

In the limit $n \rightarrow \infty$ both cases converge to a disc that can be regarded as the 2D subspace of a qubit. The extremal rays of the dual cone of polygon systems with odd number of vertices, coincide with the scaled extremal states, i.e. these systems are strongly self-dual. However, for polygon system with an even number of vertices the primal and dual cones are only isomorphic and can be matched by a rotation of $\frac{\pi}{n}$. That is, the even polygons are not strongly self-dual in the original models. We will now self-dualize these even-polygon systems using the procedure described in Theorem 3.

As discussed in section II 3 there is always the freedom to apply arbitrary bijective linear maps to all effects and the corresponding inverse map on all states. We use this to shrink the state space by $r_n \mapsto 1$ to fit in a circumscribed circle of radius one. Applying the inverse map to effects results in a effect cone with $r_n \mapsto r_n^2$. This new effect cone is strictly bigger than the cone of unnormalized states. By truncating this effect cone, such that the new extremal effect e'_i are given by

$$e'_i = \frac{1}{2} (e_i + e_{(i+1) \bmod n}) = \frac{1}{2} \begin{pmatrix} \cos\left(\frac{2\pi i}{n}\right) \\ \sin\left(\frac{2\pi i}{n}\right) \\ 1 \end{pmatrix} = \frac{\omega_i}{2}, \quad (23)$$

the primal cone coincide with the new effect cone generated by the restricted effect set.

Let us demonstrate the self-dualization procedure explicitly, by using the polygon with $n = 4$ (this is the boxworld model). In the first step the pure states and effects are transformed to the equivalent representation given in (11) and (13). In this representation the effect cone is completely embedded in the cone of unnormed states. The actual self-dualization is then done by exchanging e_i for $e'_i = \omega_i/2$, shrugging off the effects not included in the primal cone.

For all self-dualized polygon models, another interesting feature emerges for the restricted case. Namely, there exists a specific pure state $\bar{\omega}$ for each pure state ω , such that they can be perfectly distinguished by an effect e with $e(\omega) = 1$ and $e(\bar{\omega}) = 0$. Furthermore, each pair of perfectly distinguishable states can be mapped reversibly to any other pair of perfectly distinguishable states. This feature is known as *bit symmetry*, and was shown to only hold for strongly self-dual systems in the traditional framework [14].

This demonstrates that the self-dualization procedure can actually reproduce properties thought to be specific for actual strongly self-dual systems. Note that the mathematical description of actual strongly self-dual systems can be complex. Using self-dualized systems might be an alternative that helps to identify new features of strongly self-dual systems, even if one is not interested in the relaxation of the no-restriction hypothesis.

B. Spekkens's toy theory

In [15] Spekkens introduced a toy theory which replicates many features of quantum theory. For example, it exhibits a no-cloning theorem and a teleportation protocol. The theory is not explicitly probabilistic, since outcomes are not explicitly assigned probabilities. Instead, a graphical calculus is used. Given a state ω , the outcome i is only specified to be 'possible' or 'impossible'. The Spekkens theory in its original form also has no notion of arbitrary *convex mixing*, i.e. it does not have the property for any pair of states ω_1 and ω_2 , there exists a state $p\omega_1 + (1-p)\omega_2$ for all probabilities $p \in [0, 1]$.

The ability to form convex mixtures is crucial to GPTs, and in particular to its operational motivation. Fortunately, there is a natural extension of Spekkens theory which is probabilistic and which does allow convex mixing (the probabilistic version of this theory was also introduced previously by Hardy in [10]). The state space Ω of a single system is then the octahedron. In the representation that we have used, the six extremal states (i.e. the pure states) are just given by the co-ordinates of the octahedron in \mathbb{R}^3 , with an extra component for normalization. For example, the four extremal states that form the

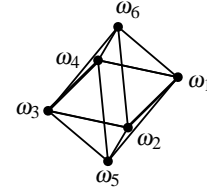


FIG. 4. The state space of the Spekkens model, with the six pure states ω_i labelled.

square base of each tetrahedron are identical to the states for boxworld (see Fig. 4). That is, for $i = 1, \dots, 4$ the states are:

$$\omega_i = \begin{pmatrix} \cos\left(\frac{2\pi i}{4}\right) \\ \sin\left(\frac{2\pi i}{4}\right) \\ 0 \\ 1 \end{pmatrix} \in \mathbb{R}^4, \quad (24)$$

and for $i = 5, 6$ the states are

$$\omega_i = \begin{pmatrix} 0 \\ 0 \\ \pm 1 \\ 1 \end{pmatrix} \in \mathbb{R}^4, \quad (25)$$

Now, the dual space of an octahedron is the cube. However, in the Spekkens theory, the space of effects is *identical* to the state space: it is also the octahedron depicted in Fig. 4. Since the octahedron can be obtained by restricting the cube (in the same way that is depicted for the hexagon in Fig. 3), we see that the Spekkens theory provides an example of a self-dualized theory. In particular, the convex probabilistic version of it is obtained using the self-dualization procedure defined in Theorem 3, and as described above for self-dualized polygons. Indeed, as with boxworld, the restricted effects are

given by:

$$e'_i = \frac{\omega_i}{2}$$

Hence we see that, at least for single systems, the Spekkens theory can be seen as an *extension* of self-dualized boxworld: the state and effect space of the Spekkens theory contain the state and effect space respectively of self-dualized boxworld. We develop the analysis of joint systems for the Spekkens theory in Section VII C 3.

We note that the single-system state space is identical to that of *stabilizer quantum mechanics*, for which the only allowed states are the eigenstates of the Pauli operators, and the allowed transformations are the Clifford operations. As discussed in [15] and further in [16], the Spekkens theory and stabilizer quantum mechanics differ in the group of reversible transformations that each theory specifies.

VI. JOINT SYSTEMS IN THE TRADITIONAL GPT FRAMEWORK

In the preceding sections we have not distinguished between single systems and joint systems. That is, our discussion so far (e.g. of self-dualization) has not involved any potential *subsystem* structure, whereby a system C can be divided into subsystems A and B , with each subsystem having well-defined states and effects. In the next section we shall consider how relaxing the no-restriction hypothesis affects composite systems. Before doing so, in this section we recall the treatment of joint systems in the traditional framework, i.e. when the no-restriction hypothesis is assumed to hold.

We will restrict the discussion of joint systems to the bipartite case with two subsystems, as the generalization of multipartite systems is straightforward. Bipartite joint states are given by elements of the product space

$$V^{AB} = V^A \otimes V^B \quad (26)$$

and joint effects are elements of $V^{AB*} = V^{A*} \otimes V^{B*}$ respectively [17].

We will represent joint states and joint effects by $n \times m$ matrices, with $n = \dim V^A = \dim V^{A*}$, $m = \dim V^B = \dim V^{B*}$. As for single systems, the application of effects on states results in the sum of the entry-wise products. This can be elegantly written as the Hilbert-Schmidt inner product

$$e^{AB}(\omega^{AB}) = \text{Tr}(e^T \cdot \omega) = \sum_{i,j} \varepsilon_{ij} \omega_{ij}, \quad (27)$$

where we write $e^T \cdot \omega$ for the matrix product between the transpose of matrix e representing the joint effect e^{AB} and the matrix ω representing the joint state ω^{AB} .

To define a composite system for a particular GPT (with specified state and effect spaces for individual systems), we must define the set of joint states $\Omega^{AB} = \{\omega^{AB}\}$, and the set of joint effects $E^{AB} = \{e^{AB}\}$, such that these are consistent with the individual systems. If the no-restriction hypothesis

holds, then, as before, once the set of joint states Ω^{AB} is defined, the set of effects E^{AB} is determined. In this situation we need only consider the definition of Ω^{AB} in order to specify the behaviour of composite systems. There is much freedom in defining Ω^{AB} , but there are two boundary cases which we now discuss.

1. Lower bound on joint systems

Consider *independently prepared* systems A and B with states $\omega^A \in \Omega^A$, $\omega^B \in \Omega^B$. Treating the systems *jointly* as a composite AB , the overall preparation is represented by the product state $\omega^{AB} = \omega^A \otimes \omega^B$, with $\omega^{AB} \in V^{AB}$. However, just as classical mixtures are allowed for single systems, for joint systems mixtures between product states give valid joint states again. This corresponds to the ability of experimenters to classically correlate the preparations and measurements of the individual systems, e.g. two experimenters can agree on specific settings.

The set of unnormalized states only containing product states and their mixtures is known as the *minimal tensor product* $A_+ \otimes_{\min} B_+$.

Definition 4. The *minimal tensor product* is given by

$$A_+ \otimes_{\min} B_+ := \left\{ \omega^{AB} \in V^{AB} \left| \begin{array}{l} \omega^{AB} = \sum_i \lambda_i \omega_i^A \otimes \omega_i^B, \\ \omega_i^A \in A_+, \omega_i^B \in B_+, \lambda_i \geq 0 \end{array} \right. \right\}. \quad (28)$$

It is the smallest possible set of unnormalized joint states ω^{AB} that is compatible with given state cones $A_+ \equiv V_+^A$, $B_+ \equiv V_+^B$ of subsystems A, B .

Similar reasoning applies to measurements, and so the set of joint effects is lower-bounded by the convex hull of product effects. Importantly, this includes the joint unit measure $u^{AB} = u^A \otimes u^B$, which is uniquely defined due to the equivalence principle. Hence, normalization of joint states ω^{AB} is represented by the condition $u^{AB}(\omega^{AB}) = 1$. This allows us to define the bipartite state space Ω_{\min}^{AB} corresponding to the minimal tensor product:

$$\Omega_{\min}^{AB} := \{ \omega^{AB} \in A_+ \otimes_{\min} B_+ \mid u^{AB}(\omega^{AB}) = 1 \} \quad (29)$$

$$= \left\{ \omega^{AB} \in V^{AB} \left| \begin{array}{l} \omega^{AB} = \sum_i p_i \omega_i^A \otimes \omega_i^B, \\ \omega_i^A \in \Omega^A, \omega_i^B \in \Omega^B, p_i \geq 0, \sum_i p_i = 1 \end{array} \right. \right\}. \quad (30)$$

For classical subsystems (i.e. a simplex), the joint states and effects defined by the minimal tensor product is sufficient to describe joint classical systems. Theories with non-classical subsystems, however, allow joint states that cannot be interpreted as a mixture of product states, i.e. entangled states. The other extreme to the minimal tensor product allows all possible entangled states, as we now show.

2. Upper bound on joint systems

Everything introduced so far is valid independent of the no-restriction hypothesis. This changes now, as we ask for the maximal sets of joint states and effects consistent with the structure of the single systems.

First, let us focus on the traditional GPT framework with single systems obeying the no-restriction hypothesis. Given a specific state space the no-restriction hypothesis determines the effects for the single systems. As argued above, the joint system should at least incorporate product effects and their mixtures. Applying such joint effects to any potential joint state ω^{AB} should give probabilities. In particular this implies that the joint states form a subset of the following set of linear elements.

Definition 5. The *maximal tensor product* is defined as

$$A_+ \otimes_{\max} B_+ := \{ \omega^{AB} \in V^{AB} \mid (e^A \otimes e^B)[\omega^{AB}] > 0, \\ \forall e^A \in E^A, e^B \in E^B \} \quad (31)$$

$$= (A_+^* \otimes_{\min} B_+^*)^*. \quad (32)$$

It is the largest possible set of unnormalized joint states ω^{AB} that is compatible with given state cones A_+, B_+ of subsystems A, B that respect the no-restriction hypothesis.

Note that the second equality arises just by definition of the dual cone (16). Hence, we see that the maximal tensor product for states is given by the maximal set of joint states consistent with the minimal tensor product for effects. Similarly the maximal tensor product for effects is defined as the maximal set of joint effects consistent with the minimal tensor product for states. Elements in the maximal tensor product, but not in the minimal tensor product are called *entangled*.

To summarise our constructions in this section: the definition of a GPT includes the tensor product, which specifies the composition of subsystems. The minimal and maximal tensor product are only the extreme cases where the joint state space Ω^{AB} is chosen as smallest or the biggest set compatible with the state spaces Ω^A, Ω^B of single systems. In general, a GPT can be defined to include any set of joint states between those extremes.

For example, the joint state space in quantum theory lies strictly between the minimal and maximal tensor product. E.g. the partial transposed of density matrices representing entangled states of two qubits or a qubit and a qutrit are known to give invalid states for the quantum tensor product, because they are not positive on all entangled effects [18]. However, these states give positive results for separable measurements, i.e. they are in the maximal tensor product. Note that these states should not be misunderstood as part of quantum theory, but form a separate toy theory that omits any entangled measurements. Nevertheless, the additional states in the maximal tensor product of local quantum systems are useful for the study of entanglement in standard quantum theory, as they correspond exactly to the set of entanglement witnesses.

3. Joint states as linear maps

For our generalization of the maximal tensor product, we shall use the following conception of joint states. Joint states can linearly map effects from one part of the joint system to unnormalized states of the other subsystem. This can be conveniently shown in the representation of joint states as matrices, since

$$(e^A \otimes e^B)[\omega^{AB}] = \text{Tr} \left[(e^A \otimes e^B)^T \cdot \omega^{AB} \right] = (e^A)^T \cdot \omega^{AB} \cdot e^B. \quad (33)$$

Using associativity of the matrix product, we can interpret parts of the expression $(e^A)^T \cdot \omega^{AB} \cdot e^B$ as ‘effective’ states of the subsystems A and B . We define these *conditional states* as

$$\omega_{e^B}^A := \omega^{AB} \cdot e^B \quad (34)$$

$$\omega_{e^A}^B := (e^A)^T \cdot \omega^{AB} \quad (35)$$

These are unnormalized states for system A and B respectively. Physically, these can be regarded as ‘post-measurement’ states on one part of the joint system, conditioned on a particular measurement outcome on the other part. This process of remotely preparing a state by a measurement on the other part of a joint state is usually referred to as ‘steering’ [19]. It demonstrates that, when measuring only part of a joint system, the joint state acts as a linear map from effects of one side of the system to unnormalized states of the other part. It can be shown that the maximal tensor product coincides exactly with all possible linear maps of this form, i.e. it corresponds to all potential joint states that have valid conditional states for non-restricted systems [2]. This property will be central for the generalization of the maximal tensor product in the next section.

Conditional states at A are unnormalized: they are weighted with the probability of obtaining the corresponding measurement outcome at B . That is, the probability accounts for the potential ignorance of the outcome for observers at B . Consequently, if one knows the measurement outcome in B the effective description of the state in A is given by the normalized conditional state:

$$\tilde{\omega}_{e^B}^A = \frac{\omega_{e^B}^A}{p(e^B | \omega^{AB})} = \frac{\omega_{e^B}^A}{u(\omega_{e^B}^A)}.$$

The *marginal state* or *reduced state* $\omega_{u^B}^A$ gives the description of the effective state on part A of a joint state ω^{AB} . This is a conditional state with $e^B = u^B$, and is already normalized i.e. $\tilde{\omega}_{u^B}^A = \omega_{u^B}^A$.

Note that this formalism still applies if the parts of the system are space-like separated, i.e. if there is no causal relationship between the measurement on the system B and the system A . However, the no-signaling principle states that steering cannot be used to transmit information, i.e. it does not allow for communication faster than the speed of light. The relationship between steering and the no-signaling principle is shown by the following theorem. First, we call a set of effects $\{e_i^A\}_i$,

for any system A , a *perfect measurement* if

$$\sum_i e_i^A = u^A.$$

An *imperfect measurement* is a set of effects $\{e_i^A\}_i$ that is not a perfect measurement.

Theorem 6. *Assuming the no-signalling principle, steering implies that all measurements are perfect measurements.*

Proof. Consider two observers in part A and B respectively sharing a joint state ω^{AB} . The observer in B performs a measurement on his part and gets some measurement outcome e_j^B . Knowing the outcome the description of the system in A from his point of view is given by the normalized conditional state $\tilde{\omega}_{e_j^B}^A$. The other observer ‘knows’ only the coarse graining of the different measurement outcomes. I.e. from his point of view the state in A is an ensemble of possible ‘post-measurement states’ $\{\tilde{\omega}_{e_j^B}^A\}$.

Remember that the equivalence principle gives a one-to-one correspondence of states and specific measurement statistics. Consequently, no-signaling requires the state in A after the measurement on B to be identical to the original marginal state $\omega_{u^B}^A$ in order to prevent information transfer, i.e.

$$\sum_i p_i \tilde{\omega}_{e_i^B}^A = \sum_i \omega_{e_i^B}^A = \omega_{\sum_i e_i^B}^A = \omega_{u^B}^A, \quad (36)$$

where we used the definition of the normalized conditional state and the linearity of effects.

Since the coarse grained conditional state needs to be equal to the marginal state for any joint state

$$\sum_i e_i^B = u^B. \quad (37)$$

□

We will use the interpretation of the maximal tensor product as the set of all positive linear maps to generalize it for systems violating the no-restriction hypothesis.

VII. THE GENERALIZED MAXIMAL TENSOR PRODUCT

As we have discussed, by removing the no-restriction hypothesis, the definition of a physical system now needs a specification of *both* the state space and the effect set. That is, the set of allowed states and the set of allowed effects can be chosen independently—except for the constraints discussed in section III. Let us now consider the specification of joint systems when the no-restriction hypothesis is removed.

The definition of the minimal tensor product $A_+ \otimes_{\min} B_+$ makes no reference to the effect sets E^A and E^B . I.e. it is constructed by products and their convex combinations. Therefore the minimal tensor product can be defined without assuming the no-restriction hypothesis, and hence carries over to our more general situation. Indeed, everything that we have introduced for joint systems so far is valid independently of the no-restriction hypothesis — with one exception.

The exception is the maximal tensor product. As before, we expect the maximal tensor product to comprise all joint states that are compatible with the given subsystems. Compatibility can be broken down to two requirements: i) non-negative results on local effects ii) valid conditional states. For non-restricted systems both requirements are equivalent, as the no-restriction hypothesis implies consistent mappings (i.e. valid conditional states) if and only if local effects give non-negative results on joint states. Now, for the general case (i.e. without the no-restriction hypothesis), valid conditional states still guarantees non-negativity on local effects. However, the implication in the other direction is no longer secured.

For example, consider attempting to use the same construction as before, i.e. we start with the minimal tensor product of effects and determine all elements of the joint system that give positive results. The resulting elements do not depend on the state spaces of the single systems *at all*, since the effects are decoupled from the state space due to the abandoned no-restriction hypothesis. Hence the resulting joint states are not forced to be consistent with the subsystems: we give an example of such a failure of consistency below.

A. Failure of the traditional maximal tensor product

Before generalizing the maximal tensor product we will show that the traditional construction rules fail for restricted systems.

The traditional maximal tensor product $A_+ \otimes_{\max} B_+$ is given by the dual of the set of separable effects. For restricted systems this yields two different variants. Equation (31) seems to suggest a construction based on the restricted effects, whereas (32) utilizes the subsystems’ dual cones, which are generated by the potential set of unrestricted effects. We show that neither choice gives the set of all joint states consistent with restricted subsystems.

The first variant is constructed as follows. Consider the restricted effects E^A of a subsystem A with an effect cone $E_+^A := \{\lambda e^A \mid e^A \in E^A, \lambda \geq 0\}$. Following equation (17) we can construct a virtual, non-restricted system \mathcal{A} with the state cone given by

$$\mathcal{A}_+ := \{\omega^A \in V^A \mid e^A(\omega^A) \geq 0 \quad \forall e^A \in E_+^A\} \supseteq A_+ \quad (38)$$

$$\Rightarrow \mathcal{A}_+^* = E_+^A. \quad (39)$$

I.e. the virtual system extends the unnormalized states, such that the no-restriction hypothesis is satisfied. Thus, the potential joint states from (31), correspond actually to the traditional maximal tensor product $\mathcal{A}_+ \otimes_{\max} \mathcal{B}_+$ of the virtual systems \mathcal{A} , \mathcal{B} .

Recall that the interpretation of joint states as positive linear maps, $\mathcal{A}_+ \otimes_{\max} \mathcal{B}_+$ is exactly the set of all maps from the restricted effect cones E_+^A (E_+^B) to the unnormalized virtual states \mathcal{B}_+ (\mathcal{A}_+) on the other side of the bipartite system. In other words, this construction includes joint states that allow the preparation of states in the subsystems not limited to the initial definition of the state spaces Ω^A , Ω^B , but to those of the virtual systems instead.

For example in a bipartite system of self-dualized boxworld with extremal states according to (11) and restricted extremal effects $e'_i = \omega_i/2$ the potential joint state

$$\omega^{AB} = \begin{pmatrix} 1 & -1 & 0 \\ 1 & 1 & 0 \\ 0 & 0 & 1 \end{pmatrix} \in \Omega_{\max}^{\mathcal{A}\mathcal{B}} \quad (40)$$

gives positive values on any pair of restricted effects. However, some conditional states are not valid for the actual system A , e.g. $\tilde{\omega}_{e'_1}^A = (-1, 1, 1)^T \notin \Omega^A$.

The second variant of the traditional maximal tensor product is based on the dual cones A_+ , B_+ according to (32). The resulting joint states are also consistent with the restricted effects, since the latter is included in the set of all of effects. However, this construction omits joint states which are consistent *only* with the restricted effects. For example, for self-dualized boxworld the identity matrix would not be included, although it has valid conditional states and gives positive results on any pair of effects.

B. Construction of the generalized maximal tensor product

As shown above, the traditional construction rules for the maximal tensor product lead to inconsistencies when applied to theories not obeying the no-restriction hypothesis. In this section we shall construct a *generalized* maximal tensor product $A_+ \otimes_{\max} B_+$: this will give the maximal set of joint states that is consistent with general subsystems, irrespective of whether the no-restriction hypothesis is assumed to hold. In other words, the generalized maximal tensor product contains all bipartite states whose conditional (i.e. also marginal) states are elements of the original state spaces.

Definition 7. The *generalized maximal tensor product* of systems A , B with primal cones A_+ , B_+ , dual cones A_+^* , B_+^* and effect cones E_+^A , E_+^B is given by

$$\begin{aligned} A_+ \otimes_{\max} B_+ &:= (E_+^A \otimes_{\min} B_+^*)^* \cap (A_+^* \otimes_{\min} E_+^B)^* \\ &= (E_+^A \otimes_{\min} B_+^* \cup A_+^* \otimes_{\min} E_+^B)^* \end{aligned} \quad (41)$$

For the saturated case, dual cones and effect cones are identical, and we recover the usual maximal tensor product as follows.

Proposition 8. Suppose that $E_+^A = A_+^*$ and $E_+^B = B_+^*$. Then $A_+ \otimes_{\max} B_+ = A_+ \otimes_{\max} B_+$.

Proof. Under the assumptions, Eq. 41 becomes

$$\begin{aligned} A_+ \otimes_{\max} B_+ &= (A_+^* \otimes_{\min} B_+^*)^* \\ &= A_+ \otimes_{\max} B_+ \end{aligned}$$

using the definition of the maximal tensor product in (31). \square

Hence our construction is indeed a generalization of the existing definition of the maximal tensor product. It determines all joint states consistent with general subsystems regardless

whether the no-restriction hypothesis holds or not. I.e. all joint states with valid conditional states are included, as shown in the following theorem.

Theorem 9. Let $\omega^{AB} \in V^{AB}$. Then $\omega^{AB} \in A_+ \otimes_{\max} B_+$ iff ω^{AB} has well-defined conditional states:

$$\tilde{\omega}_{e_B^A}^A \in \Omega^A \text{ and } \tilde{\omega}_{e_A^B}^B \in \Omega^B$$

for all $e^A \in E^A$ and $e^B \in E^B$.

Proof. We shall show that

$$\omega_{e_B^A}^A \in A_+ \text{ iff } \omega^{AB} \in (E_+^A \otimes_{\min} B_+^*)^* \quad (42)$$

and that

$$\omega_{e_A^B}^B \in B_+ \text{ iff } \omega^{AB} \in (A_+^* \otimes_{\min} E_+^B)^*. \quad (43)$$

Since $A_+ \otimes_{\max} B_+$ is defined as the intersection of sets of linear maps $(E_+^A \otimes_{\min} B_+^*)^*$ and $(A_+^* \otimes_{\min} E_+^B)^*$, this will establish the thesis.

First we show the $A \rightarrow B$ direction, i.e. (42), which is the statement that $(E_+^A \otimes_{\min} B_+^*)^*$ is the set of all and only those joint states ω^{AB} such that each ω^{AB} defines a map from effects $e^A \in E_+^A$ on system A to valid unnormalized states $\omega^B \in B_+$.

Recall that for non-restricted systems the traditional maximal tensor product is already known to give all positive linear maps from the effect cone of one part of the system to the state cone of the other part for both directions. We now show that $(E_+^A \otimes_{\min} B_+^*)^*$ can be interpreted as the traditional maximal tensor product $\mathcal{A}_+ \otimes_{\max} \mathcal{B}_+$ of two virtual systems \mathcal{A} and \mathcal{B} that obey the no-restriction hypothesis. \mathcal{A} is the virtual system that has already been introduced in (38). It has an extended virtual state cone \mathcal{A}_+ , since $(E_+^A)^* \subset \mathcal{A}_+$. However, the actual effect cone is kept, as it coincides with the dual cone $E_+^A = \mathcal{A}_+^*$ characterizing unnormalized effects of the non-restricted systems. The opposite situation applies to \mathcal{B} . Here, the effect set E_+^B is extended to the dual cone B_+^* , so that $E_+^B \subset B_+^*$, where B_+^* representing the full set of potential unnormalized effects. However, the original state cone $B_+ = \mathcal{B}_+$ is kept. With these conventions

$$(E_+^A \otimes_{\min} B_+^*)^* = \mathcal{A}_+ \otimes_{\max} \mathcal{B}_+$$

follows directly from the definition of the traditional tensor product in (32). Hence for the $A \rightarrow B$ direction, this means that $(E_+^A \otimes_{\min} B_+^*)^*$ contains all positive linear maps from the restricted effects in A to allowed states in B . That is, $\omega^{AB} \in (E_+^A \otimes_{\min} B_+^*)^*$ is a sufficient and necessary condition for $\omega_{e^A}^B \in B_+$ which proves (42). However, for the traditional maximal tensor product the same joint states also coincide with the positive maps in the other direction $B \rightarrow A$, potentially including invalid mappings.

By swapping the roles of A and B in the above argument, we similarly obtain that the set $(A_+^* \otimes_{\min} E_+^B)^*$ includes all linear maps that are consistent for the $B \rightarrow A$ direction, but also those which lead to inconsistencies in $A \rightarrow B$ opposite direction. Hence we obtain (43). \square

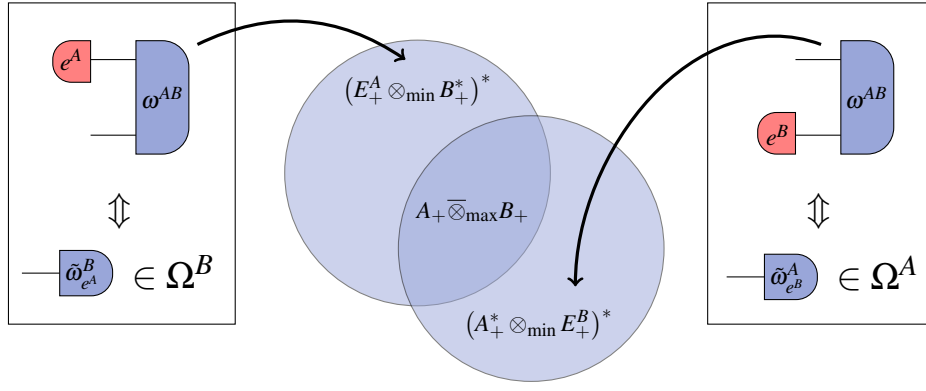


FIG. 5. Illustration of the construction of the generalized maximal tensor product

Theorem 9 shows that the generalized maximal tensor product includes only those joint states that are consistent in both directions (i.e. the intersection of the sets $(E_+^A \otimes_{\min} B_+^*)^*$ and $(A_+^* \otimes_{\min} E_+^B)^*$). Note that if ω^{AB} has well-defined conditional states, then in particular it is locally positive:

$$(e^A \otimes e^B) [\omega^{AB}] \geq 0$$

which provides a useful necessary condition that joint states must satisfy.

In the traditional GPT framework the choices of tensor products for states and effects are not independent, as the no-restriction hypothesis does not only apply to single systems, but to the joint system as well. Having the minimal tensor product for joint states (effects) does in fact constitute the maximal tensor product for the set of joint effects (states). This restriction seems inappropriate given that arbitrary single systems can actually be emulated by classical systems with constrained measurements [20], whereas entanglement is a strictly non-classical feature.

In our modified framework that is also valid for systems violating the no-restriction hypothesis, this is no longer the case. We have seen that we can generalize the maximal tensor product, but nevertheless we are not forced to use this for states when we choose the minimum tensor product for effects and the other way round.

C. Examples of joint systems

To give some specific examples for the generalized maximal tensor product, we have calculated it for the toy theories introduced in sections IV and V using the double description method [21].

1. Noisy boxworld

In the original unrestricted version of boxworld joint systems are given by the maximal tensor product, including the 16 extremal product states and 8 pure entangled joint states

$\Phi^{AB} = \frac{1}{2}(\omega_1 \otimes \omega_2 - \omega_2 \otimes \omega_1 + \omega_2 \otimes \omega_3 + \omega_3 \otimes \omega_1)$ and respectively the states transformed by local symmetries.

These entangled extremals can be interpreted as a maximally entangled state of two such systems, as they form an isomorphic map and have totally mixed reduced states. They correspond to a rotation of $\frac{\pi}{4}$ and the local symmetries of the state spaces.

This theory has become very popular as it shows nonlocal correlations beyond those possible in quantum theory, when choosing between two possible binary measurements at each side of the bipartite systems. Let us denote the two measurements $\{M_x^A\}$ and $\{M_y^B\}$ for each of the systems A and B respectively: we index the measurements at each system with $x, y \in \{0, 1\}$. Each measurement has binary outcomes, labelled with $a, b \in \{0, 1\}$ for systems A and B respectively. For example, the $x = 0$ measurement on system A consists of a pair of effects $M_x^A = \{e_0, e_1\}$ satisfying $e_0 + e_1 = u$; similarly for the $x = 1$ measurement on system A , and $y \in \{0, 1\}$ measurements on system B . This leads to a bipartite conditional probability distribution

$$P(a, b|x, y) := (e_a \otimes e_b)[\omega^{AB}] \quad (44)$$

We define the correlation

$$C_{xy} := P(a = b|x, y) - P(a \neq b|x, y).$$

To introduce the Clauser-Horne-Shimony-Holt (CHSH) inequality for demonstrating nonlocality, we introduce the parameter

$$S := |C_{00} + C_{01} + C_{10} - C_{11}|,$$

For classical systems it is upper bounded by the CHSH inequality [22]

$$S^C \leq 2,$$

whereas for quantum theory it must satisfy $S^Q \leq 2\sqrt{2}$ [23]. However, local measurements on the maximally entangled state Φ in boxworld can produce correlations which reach the algebraic maximum $S^{\max} = 4$, i.e. the theory allows the post-quantum correlations known as PR boxes [13].

For the noisy version of boxworld introduced in section IV there is still a notion of a maximally entangled state in the generalized maximal tensor product, namely

$$\Phi^\lambda = \omega_{\text{ent},1}^{AB} = \text{diag}\left(\frac{1}{\lambda}, \frac{1}{\lambda}, 1\right) \cdot \Phi^1, \quad (45)$$

i.e. the original maximally entangled state Φ^1 combined with a mapping of the effects on one side of the system to the original unrestricted set. Note that this map does not undo the restriction of effects completely. The reversion only happens to occur in this particular case when mapping to states of the other part. On the other part, however, only restricted effects can be applied to. Consequently, the correlations possible with restricted systems will be different to those possible in unrestricted systems.

Furthermore, constructing the generalized maximal tensor product it turns out there are 4 different classes of new pure joint states that are entangled but not maximally entangled. These are representatives of each class

$$\begin{aligned} \omega_{\text{ent},2}^{AB} &= -\alpha \omega_2 \otimes \omega_2 + \beta \omega_2 \otimes \omega_4 + \beta \omega_4 \otimes \omega_2 - \alpha \omega_4 \otimes \omega_4 \\ \omega_{\text{ent},3}^{AB} &= -\alpha \omega_2 \otimes \omega_2 + \beta \omega_2 \otimes \omega_3 + \beta \omega_4 \otimes \omega_2 - \alpha \omega_4 \otimes \omega_3 \\ \omega_{\text{ent},4}^{AB} &= -\alpha \omega_2 \otimes \omega_2 + \beta \omega_2 \otimes \omega_4 + \beta \omega_3 \otimes \omega_2 - \alpha \omega_3 \otimes \omega_4 \\ \omega_{\text{ent},5}^{AB} &= -\alpha \omega_3 \otimes \omega_2 + \beta \omega_4 \otimes \omega_1 + \beta \omega_4 \otimes \omega_2 - \alpha \omega_4 \otimes \omega_3 \end{aligned}$$

with $\alpha = \frac{1-\lambda}{4\lambda}, \beta = \frac{1+\lambda}{4\lambda},$ (46)

where the other elements of the class only differ by the local symmetries.

In conclusion the generalized maximal tensor product is spanned by 96 pure states. Namely, it consists of 16 local pure states, 8 pure entangled states of class $\omega_{\text{ent},1}^{AB}$, 8 of class $\omega_{\text{ent},2}^{AB}$, 16 of class $\omega_{\text{ent},3}^{AB}$, 16 of class $\omega_{\text{ent},4}^{AB}$ and 32 states of class $\omega_{\text{ent},5}^{AB}$.

Considering local measurements on one instance of any of the nonlocal extremal states the maximal CHSH violation S^λ as a function of the parameter λ of the restricted model can be shown to be $4\lambda^2$. Note that this bound is only guaranteed for the correlations that occur from direct measurements. However, it is known that wiring the measurements on multiple joint states via classical post-processing, might give rise to a distillation of correlations beyond for some values of λ [24].

2. Self-dualized polygons

Interestingly, not only boxworld but all bipartite polygon systems allow a joint state with features known from the maximally entangled state of ordinary quantum theory. Namely, the linear maps corresponding to these states are given by isomorphisms of the dual and primal cones with maximally mixed reduced states. The $2n$ different maximally entangled states correspond to the elements of the dihedral group. For even n , the maximally entangled states include an additional rotation of π/n mapping the dual cone of one part to the primal cone of the other part. It was shown that non-local correlations based on two binary local measurements

on the maximally entangled states at each side show correlations strictly weaker than quantum correlations for the odd case, whereas the unrestricted even case shows correlations as strong as those of quantum theory or stronger [9].

Replacing the original polygon systems with even n by their self-dualized versions, the maximally entangled states lose the additional rotation as the new effect cone and the state cone coincide. Note, that the self-dualized single systems become subtheories of the theory given in the limit $n \rightarrow \infty$, i.e. the quantum case, as both states and effects form strict subsets. Thus, the correlations on the maximally entangled state form a strict subset of those in quantum theory, in contrast to the unrestricted case which allows post-quantum correlations. Even though the restricted polygons are not genuine strongly self-dual but only self-dualized, this is consistent with the conjecture in [9], that strong self-duality limits correlations.

For self-dualized boxworld the generalized maximal tensor product is given by the 16 local pure states, the 8 states $\omega_{\text{ent},1}^{AB}$ representing the identity and symmetry mappings as well as a class of 64 pure entangled states $\omega_{\text{ent},2}^{AB} = 1/4(-\omega_1 \otimes \omega_1 + \omega_1 \otimes \omega_3 + 2\omega_2 \otimes \omega_4 + \omega_3 \otimes \omega_1 - \omega_3 \otimes \omega_3 + 2\omega_4 \otimes \omega_2)$.

Unfortunately, using the double description method, we were not able to characterize all extremals of the generalized maximal tensor product for polygon systems with a higher number of vertices.

3. Spekkens's toy theory

The Spekkens theory that we introduced earlier is a local theory, meaning that (in the probabilistic version) it cannot violate any Bell inequalities. However, as discussed in [15], the Spekkens theory has entangled states. This raises the question of why the Spekkens theory does not exhibit bipartite nonlocality. In contrast, a classical theory, i.e. a simplex, is local but it does not have entangled states. One could then ask, given that the Spekkens theory has entangled states, but is local, what must be *added* to the definition of the theory to make it nonlocal?

In our framework, the answer to this question can be clearly understood in terms of the geometry of the state space. First, recall that the state space Ω^A of a single system in the Spekkens theory is an octahedron, and the effect space E^A is identically the same, i.e. E^A not the full dual space. Consider a pair of single systems A and B in the Spekkens theory. Since the effect space E^A is not the full dual space A_+^* , we must use the generalized tensor product $\Omega^{AB} = A_+ \otimes_{\max} B_+$ to define the bipartite states. Then consider the following bipartite state:

$$\omega^{AB} = \begin{pmatrix} 0 & 0 & 0 & 0 \\ 0 & -\frac{1}{2} & -\frac{1}{2} & 0 \\ 0 & -\frac{1}{2} & \frac{1}{2} & 0 \\ 0 & 0 & 0 & 1 \end{pmatrix} \quad (47)$$

It is straightforward to verify that ω^{AB} leads to well-defined conditional states for system B for all effects $e^A \in E^A$, i.e.:

$$\tilde{\omega}_{e^A}^B \in \Omega^B$$

and correspondingly for conditional states for system A when using effects on system B . In particular, it is also easily checked that $(e^A \otimes e^B)[\omega^{AB}] \geq 0$ for any pair of effects e^A and e^B . Hence by Theorem 9, this shows that ω^{AB} is in the generalized tensor product $A_+ \overline{\otimes}_{\max} B_+$ for the Spekkens theory.

Now, since the Spekkens theory is local, the CHSH inequality (VII C 1) is satisfied for any choice of measurements M_x and M_y on the state ω^{AB} , or any other bipartite state. However, let us consider the unrestricted effect space A_+^* from which the restricted space E^A for the Spekkens theory was derived. The unrestricted effect space of the octahedron is the cube. We can represent the normalised extremal effects as the vertices of a cube:

$$e_i = \frac{1}{2} \begin{pmatrix} \pm 1 \\ \pm 1 \\ \pm 1 \\ 1 \end{pmatrix}$$

Now, suppose that we use the cube to be the effect space for the octahedron, i.e. we use the full dual space. It is easily shown that the state ω^{AB} defined in Eq. 47 is again in the generalized maximal tensor product $A_+ \overline{\otimes}_{\max} B_+$. However, we can now provide measurements which violate the CHSH inequality. In particular, consider two measurements for Alice given by $M_0^A = \{e_0, u - e_0\}$ and $M_1^A = \{e_1, u - e_1\}$ where:

$$e_0 = \frac{1}{2} \begin{pmatrix} 1 \\ 1 \\ -1 \\ 1 \end{pmatrix}, e_1 = \frac{1}{2} \begin{pmatrix} -1 \\ -1 \\ -1 \\ 1 \end{pmatrix}$$

and two measurements for Bob given by $M_0^B = \{e_0, u - e_0\}$ and $M_1^B = \{e_2, u - e_2\}$, where:

$$e_0 = \frac{1}{2} \begin{pmatrix} 1 \\ 1 \\ -1 \\ 1 \end{pmatrix}, e_2 = \frac{1}{2} \begin{pmatrix} -1 \\ -1 \\ 1 \\ 1 \end{pmatrix}$$

By using these choices of measurements in Eq. 44 and the following equations, we obtain the value of the CHSH parameter: this is $S = 4$. This is the value attained by PR boxes, and hence using the full effect space essentially yields the same nonlocality as boxworld.

We therefore see that the Spekkens theory can be embedded into a nonlocal theory by embedding the effect space of single system into the full dual cone. Moreover, we see completing the Spekkens theory in this way yields boxworld. This provides a new understanding of why the Spekkens theory is local: the measurements are too restricted.

VIII. CONCLUSIONS

We have extended the framework of generalized probabilistic theories. Given an arbitrary state space the traditional framework determines the possible measurement outcomes as corresponding to the complete set of probability valued linear functionals on states. In contrast to the traditional framework, our generalization allows the set of states and the set of effects to be defined separately. As a result the upper bound for the set of joint states, known as the maximal tensor product, is no longer valid in its traditional form, but has to be replaced by a generalized version.

As an application for restricted models, we provided a self-dualization procedure that alters any theory by restricting the set of effects, such that states and the restricted effects are similarly related as states and unrestricted effects in strongly self-dual systems. We introduce specific examples for which the self-dualization does not only give a formal resemblance but reproduces a phenomenon called *bit symmetry* shown to only hold for strongly self-dual systems in the traditional framework [14]. Furthermore, these self-dualized models show quantum correlations, whereas the original models have correlations that are stronger than quantum correlations. In particular, the correlations of boxworld—a theory known to allow correlations only restricted by the no-signalling principle—has classical correlations if self-dualized, even though the generalized maximal tensor product includes maximally entangled states. We showed how the Spekkens theory is related to this model, since it is also self-dual and violates the no-restriction hypothesis: but were it to satisfy this principle, by taking the full dual cone, it would produce nonlocal correlations.

As another application for restricted models, we show that restrictions can be used to alter theories, such that their measurements are inherently noisy. This is different to the unrestricted theories, since in our noisy theories it holds that for pure states there is no non-trivial extremal effect occurring with certainty. We derive the maximal CHSH violation [22] of a noisy version of boxworld as a function of a noise parameter.

The modified framework is therefore suitable for examining new situations that could not be addressed using the traditional framework. In particular the self-dualization procedure might be useful for the study of strong self-duality that has recently received much interest [2, 9, 14].

ACKNOWLEDGMENTS

We thank Jonathan Barrett, Christian Gogolin and Haye Hinrichsen for insightful discussions. PJ is supported by the German Research Foundation (DFG). RL is supported by the Templeton Foundation.

[1] J. Barrett, Phys. Rev. A **75**, 032304 (2007), URL <http://link.aps.org/doi/10.1103/PhysRevA.75.032304>.

[2] H. Barnum, J. Barrett, M. Leifer, and A. Wilce, preprint (2008),

- 0805.3553v1.
- [3] V. Scarani, N. Gisin, N. Brunner, L. Masanes, S. Pino, and A. Acin, *Phys. Rev. A* **74**, 042339 (2006), URL <http://link.aps.org/doi/10.1103/PhysRevA.74.042339>.
 - [4] L. Masanes and M. P. Müller, *New Journal of Physics* **13**, 063001 (2011), URL <http://stacks.iop.org/1367-2630/13/i=6/a=063001>.
 - [5] L. Hardy, preprint (2011), 1104.2066.
 - [6] G. Chiribella, G. M. D'Ariano, and P. Perinotti, *Phys. Rev. A* **84**, 012311 (2011), URL <http://link.aps.org/doi/10.1103/PhysRevA.84.012311>.
 - [7] A. Short and J. Barrett, *New Journal of Physics* **12**, 033034 (2010).
 - [8] H. Barnum, J. Barrett, L. O. Clark, M. Leifer, R. Spekkens, N. Stepanik, A. Wilce, and R. Wilke, *New Journal of Physics* **12**, 033024 (2010), URL <http://stacks.iop.org/1367-2630/12/i=3/a=033024>.
 - [9] P. Janotta, C. Gogolin, J. Barrett, and N. Brunner, *New Journal of Physics* **13**, 063024 (2011), URL <http://stacks.iop.org/1367-2630/13/i=6/a=063024>.
 - [10] L. Hardy, preprint (1999), quant-ph/9906123.
 - [11] I. Bengtsson, S. Weis, and K. Życzkowski, preprint (2011), 1112.2347.
 - [12] G. Ver Steeg and S. Wehner, *QIC* **9**, 0801 (2009).
 - [13] S. Popescu and D. Rohrlich, *Foundations of Physics* **24**, 379 (1994), ISSN 0015-9018, URL <http://dx.doi.org/10.1007/BF02058098>.
 - [14] M. P. Müller and C. Ududec, *Phys. Rev. Lett.* **108**, 130401 (2012), URL <http://link.aps.org/doi/10.1103/PhysRevLett.108.130401>.
 - [15] R. W. Spekkens, *Phys. Rev. A* **75**, 032110 (2007), URL <http://link.aps.org/doi/10.1103/PhysRevA.75.032110>.
 - [16] B. Coecke, B. Edwards, and R. Spekkens, *Electronic Notes in Theoretical Computer Science* **270**, 15 (2011).
 - [17] It can be shown that this follows from two conditions on joint states: i) local tomography ii) the no-signalling principle. The no-signalling principle forbids to send information by local operations on a joint state and will be explained in more detail in section VI 3. Local tomography is the identification of joint states by combinations of local measurements.
 - [18] M. Horodecki, P. Horodecki, and R. Horodecki, *Physics Letters A* **223**, 1 (1996), ISSN 0375-9601, URL <http://www.sciencedirect.com/science/article/pii/S0375960196007062>.
 - [19] J. Oppenheim and S. Wehner, *Science* **330**, 1072 (2010), URL dx.doi.org/10.1126/science.1192065.
 - [20] A. S. Holevo, *Probabilistic and Statistical Aspects of Quantum Theory*, vol. 1 of *North-Holland Series in Statistics and Probability* (North-Holland, Amsterdam, 1982).
 - [21] K. Fukuda and A. Prodon, in *Combinatorics and Computer Science*, edited by M. Deza, R. Euler, and I. Manoussakis (Springer, 1996), vol. 1120 of *Lecture Notes in Computer Science*, pp. 91–111.
 - [22] J. F. Clauser, M. A. Horne, A. Shimony, and R. A. Holt, *Phys. Rev. Lett.* **23**, 880 (1969), URL <http://link.aps.org/doi/10.1103/PhysRevLett.23.880>.
 - [23] B. Cirel'son, *Letters in Mathematical Physics* **4**, 93 (1980), ISSN 0377-9017, URL <http://dx.doi.org/10.1007/BF00417500>.
 - [24] J. Allcock, N. Brunner, N. Linden, S. Popescu, P. Skrzypczyk, and T. Vértesi, *Phys. Rev. A* **80**, 062107 (2009), URL <http://link.aps.org/doi/10.1103/PhysRevA.80.062107>.